

Graduiertenkolloquium Angewandte Informatik

Large-Scale Pattern-Based Information Extraction from the World Wide Web

M. Sc. Sebastian Blohm

blohm@kit.edu
Institut AIFB

Suppose you want to stay up-to-date with all relevant facts about a product on the market or quickly generate a list of interesting events at your next travel destination? This data is probably nowhere covered broader and up-to-date than on the Web. Wouldn't it be convenient to simply aggregate the required facts from the Web?

Extracting information from text is the task of obtaining structured, machine-processable facts from information that is mentioned in an unstructured manner. It thus allows systems to automatically aggregate information for further analysis, efficient retrieval, automatic validation, or appropriate visualization. Information Extraction systems require a model that describes how to identify relevant target information in texts. These models need to be adapted to the exact nature of the target information and to the nature of the textual input, which is typically accomplished by means of Machine Learning techniques that generate such models based on examples.

The talk presents studies on using textual patterns for extracting information from the World Wide Web. We explore both fundamental design choices and practical applications.

The talk will be given in German language.

Termin: Mittwoch, 18. November 2009, 15:45 Uhr

Ort: Englerstraße 11, 76131 Karlsruhe
Kollegiengebäude am Ehrenhof (Geb. 11.40), 2. OG, Raum 231
(Hinweise für Besucher: www.aifb.uni-karlsruhe.de/Allgemeines/Besucher)

Veranstalter: Institut AIFB, Forschungsgruppe Wissensmanagement

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

Andreas Oberweis, Hartmut Schreck, Detlef Seese, Wolffried Stucky, Rudi Studer (Org.), Stefan Tai